EXHIBIT A

UNITED STATES DISTRICT COURT

FOR THE NORTHERN DISTRICT OF CALIFORNIA

SAN FRANCISCO DIVISION

RICHARD KADREY, et al., *Individual and Representative Plaintiffs*, v. META PLATFORMS, INC., *Defendant.*	Case No. 3:23-cv-03417-VC

REBUTTAL REPORT OF DAVID R. CHOFFNES, Ph.D.

FEBRUARY 26, 2025

Table of Contents

Introduction

1. My name is David R. Choffnes. I am currently an Associate Professor in Computer Science and Executive Director of the Cybersecurity and Privacy Institute at Northeastern University in Boston, MA. I have been a professor of computer science for more than 11 years and Executive Director for nearly four years. For approximately the last five years, I have been providing consulting services to the legal industry. These consulting services primarily involve the analysis of software systems, to understand how they operate and what data they share. I have been an active researcher in the field of computer science for the past 21 years, and I am an expert in topics that include computer networking and distributed systems. I have a Ph.D. in Computer Science from Northwestern University (2010) and am a Senior Member of the Association of Computing Machinery (ACM) since 2021. My Ph.D. thesis focused on building computer systems on top of BitTorrent. My expertise includes the analysis of peer-to-peer (P2P) file-sharing technologies, such as BitTorrent, and their implications in legal contexts.

2. My curriculum vitae (CV), which contains all publications I have authored in the past 10 years, is attached hereto as Appendix A. In the past four years, I have testified as an expert witness in one confidential arbitration and have not provided deposition or trial testimony.

3. I am being compensated at a rate of $600 per hour for my work on this matter and $1,000 per hour for any deposition testimony. My compensation is not contingent or dependent on my testimony, on the content of this report, or on the outcome of this case in any way.

4. A list of the materials that I have considered in rendering my opinions for this report is listed in Appendix B. I reserve the right to offer additional opinions based on any additional discovery or additional assignments with which I am provided.

5. **Summary of Key Points**: This rebuttal will focus on three main areas: affirmative evidence of torrenting, misleading statements in Barbara Frederiksen-Cross's report, and missing evidence that could identify the extent of Meta's torrenting, including leeching and seeding. Given the evidence provided and the particularly the limited nature of the evidence in certain areas, there is a high likelihood that additional relevant evidence would illuminate activities such as leeching, the use of torrents for Libgen.rs fiction, and sharing Plaintiffs' data with other BitTorrent peers.

Affirmative Evidence of Torrenting

6. **Explicit Admissions in Frederiksen-Cross's Report.** Frederiksen-Cross's report contains several explicit admissions of torrenting activities by Meta. This admission is crucial as it directly contradicts her overall assertion that Meta did not engage in large-scale sharing of Plaintiffs' works given what we know about how BitTorrent works.

7. Specifically, Frederiksen-Cross's report explicitly confirms that Meta used BitTorrent to download and share all of Plaintiffs' works.[1] In more detail, the report admits to using BitTorrent to download and share 666 copies of Plaintiffs' works across 194 torrents. As discussed below, there is circumstantial evidence that the extent of torrenting may go beyond these 666 copies of Plaintiffs' works. In addition, Frederiksen-Cross's report further admits the use of BitTorrent in general to download and share Plaintiffs' works.[2]

Misleading Statements and Omissions in Frederiksen-Cross's Report

8. **Network Configuration and Uploading Capabilities.** Frederiksen-Cross's claim about network configurations preventing uploading is misleading. She states, "Meta's network configurations would have blocked any connections not initiated by Meta." However, nearly every home router firewall blocks unsolicited inbound connections. The fact that Meta's firewall worked this way was not unique. It instead was standard.

9. Additionally, BitTorrent is designed to work around this configuration. Note that BitTorrent facilitates data exchanges between peers, and works best when peers are sharing data with each other in a reciprocal way (i.e., uploading data to a peer from which it is downloading data). Given this context, there are two entirely common ways that BitTorrent would engage in uploading during the entire leeching and seeding process (i.e., during the download and after the download is complete).

10. First, as admitted in the Frederiksen-Cross report, the Meta BitTorrent client may initiate connections to other leechers while the download is not complete, for the purpose of fetching pieces of torrent data from those peers. While doing so, BitTorrent is designed to provide pieces of torrent data to those leechers (i.e., upload to them) to increase the chances that the peer will continue to provide pieces for downloading.

11. Second, the libtorrent client that Meta used can enable connections that circumvent the network configuration's network blocking, via a technique called "hole punching." Specifically, libtorrent implements a Peer Exchange protocol that attempts to allow incoming connections by first attempting an outbound connection.[3] As the author of libtorrent states, "the main two approaches used by libtorrent (and bittorrent clients generally) are: commonly connected peers may introduce two NATed peers to each other via the peer exchange extension. In this mode both peers try to connect to each other simultaneously, hoping that both NATs will open up pin-holes for the ports that are being attempted." The term "NAT" refers to "network address translation" and "pin-holes" refers to hole punching as described above. The NAT configuration referenced in the above quote is essentially identical (from the perspective of how unsolicited inbound network connections are blocked, but ones initiated by Meta are not) to the one used in Meta's EC2 instances used for torrenting. More

[1] Frederiksen-Cross Report ¶94, ¶110, and Table 2.

[2] Frederiksen-Cross Report ¶80 and footnote 199.

[3] https://stackoverflow.com/questions/66377090/libtorrent-nat-traversal#66476983

details about how libtorrent opens such "pin-holes" in the network firewall are detailed in the BEP 55 standard.[4]

12. In the Frederiksen-Cross report, neither of the above uploading mechanisms was tested in the EC2 networking environment where Plaintiffs' works were torrented. In other words, Frederiksen-Cross did not test whether Meta's libtorrent client uploads data to peers when this client initiates peer connections. Nor did Frederiksen-Cross test whether "hole punching" allowed new peer connections that led to Meta's libtorrent client uploading pieces of torrent data. Instead, her only experiment involved a simple "netcat" (or "nc") command issued on BitTorrent's default port, which yielded expected results but ones that are *irrelevant to the behavior of libtorrent*.

13. In contrast, the above facts strongly suggest that Meta shared torrented content (including the Plaintiffs' works) while downloading and seeding over 200 terabytes of data via BitTorrent.

14. Further, the Frederiksen-Cross report does not discuss documents where Meta employees indicated they wanted to keep torrenting activity off of Meta infrastructure so that seeders could not be traced back to Meta IP addresses.[5] Such statements indicate knowledge of how BitTorrent shares content while downloading and seeding, and, relatedly, that employees seemed to know this conduct was not appropriate.[6] In other words, Meta employees seemed to be quite aware that their use of BitTorrent would cause Plaintiffs' works to be reshared by Meta, in contrast to the direct download attempts they had previously tried.

15. **Focus on Byte Percentages Instead of Piece Percentages.** Frederiksen-Cross's report places unnecessary emphasis on Plaintiffs' works being a small portion of each torrent when counting the number of total bytes associated with the torrent file.[7] *This is irrelevant*. BitTorrent uploads and downloads data in fixed-size chunks called pieces and does not distinguish the content associated with each piece *at all*. To BitTorrent, the sizes of files associated with each torrent do not affect uploading at all. The only thing that matters is which pieces are available to download and which ones are needed. As such, the emphasis on bytes and file sizes is misleading. Further, the percentage of pieces containing Plaintiffs' works is higher compared to the byte count, which Frederiksen-Cross's report downplays.

16. **Speculative Statements about Distribution Prevention.** Frederiksen-Cross makes speculative claims about preventing any distribution of Plaintiffs' works. She states in paragraph 100, "In sum, Meta took steps to prevent seeding data downloaded via BitTorrent, and these steps should have prevented any distribution of Plaintiffs' works by Meta." This

---

[4] https://www.bittorrent.org/beps/bep_0055.html

[5] Meta_Kadrey_00108327, Meta_Kadrey_00235448.

[6] Meta_Kadrey_00232764, Meta_Kadrey_00235868, Meta_Kadrey_00235739.

[7] She states, "The small proportion of Plaintiffs' works to the sizes of the at-issue datasets, as well as the small proportion of Plaintiffs' work pieces within the torrents they exist in, are important factors that further drive down the possibility that these works would have been seeded to other peers on the network." Frederiksen-Cross Report ¶121.

statement is speculative and not supported by concrete evidence. She ignores that uploading occurs during leeching (downloading) and seeding phases; she ignores her own statement that reciprocal uploading can happen when Meta's BitTorrent client establishes a connection to a leecher regardless of the network configuration that prevents unsolicited inbound connections; and she does not provide any evidence that the Meta's use of BitTorrent actually prevented any sharing of the Plaintiffs' works.

17. **Ignoring BitTorrent's Functionality with Multiple Peers.** Frederiksen-Cross's analysis ignores how BitTorrent works with multiple peers downloading and uploading simultaneously. In paragraphs 64 and 65, she refers to the downloading phase without any reference to the fact that all BitTorrent peers (including Meta's) are designed to upload torrent pieces to peers that are connected, interested in pieces, and unchoked. Recall that an unchoked peer is one that is selected to receive data (pieces of the torrent) already downloaded by BitTorrent. In libtorrent, the BitTorrent client used by Meta, up to 8 peers can be unchoked (i.e., uploaded to) at a time by default. Further, the report overlooks the fact that BitTorrent's efficiency and speed comes from its ability to download/upload pieces of a torrent to/from multiple peers simultaneously. As I describe below, these uploading behaviors, when considered across multiple connected peers simultaneously, leads to a high likelihood that Meta shared at least one (if not more) of Plaintiffs' works.

18. **Opportunities for Meta to Upload to Peers During Leeching Phase.** The Frederiksen-Cross report mentions that BitTorrent downloads the rarest pieces of torrent data first, but it does not recognize that the rarest pieces are necessarily the most attractive pieces of data for any other leechers in the swarm. As a result, when Meta downloaded rare pieces, they were the most likely to be uploaded by Meta to other leechers during the download phase.

19. **Omission of leeching duration.** Frederiksen-Cross's report claims that Meta's safeguards rendered the possibility of seeding highly unlikely, which is misleading. She states, "Meta's torrent download scripts limited the amount of time a torrent could theoretically have been seeding." This refers to the fact that Meta's use of BitTorrent limited *seeding* to at most sixty seconds. However, this does not account for the *leeching* phase, where BitTorrent both downloads from and uploads to other peers, and is the dominant factor in Meta's sharing of Plaintiffs' works. The torrents that Meta downloaded consisted of billions of bytes of data, likely taking multiple hours of leeching before seeding began. In fact, evidence in Meta's developer notes indicates that torrenting may have been going on for *weeks*.[8] Meta did not furnish data about when each download started and ended, thus hiding exactly how much opportunity there was for sharing data during the leeching period. As pointed out in the Frederiksen-Cross report, BitTorrent frequently changes the set of peers it uploads to and downloads from ("choking" and "unchoking" every 15 seconds), and thus during the leeching phase there were likely thousands if not tens of thousands of opportunities to share Plaintiffs' works with other peers.

[8] Meta_Kadrey_00088043, where Xiaolan Wang's update on April 1, 2024 mentioned "Anna's Archive" in "Plan for this week", April 8 update mentioned "Anna's archive Anna's Archive Logbook: - Kick off data downloading (still ongoing)," April 15 update mentioned "AA downloading continued" and April 22 reports most but not all data is downloaded.

20. **Omission of statistics to determine likelihood of piece sharing.** In the Frederiksen-Cross report, she states that it is "exceedingly unlikely" that Meta shared Plaintiffs' works but provides no statistical analysis whatsoever to support that claim. In contrast, I conducted a statistical analysis on the data considered in the report. Table 1 provides a summary of my analysis, with the first column representing the datasets mentioned in Table 4 of Frederiksen-Cross's report, and the second column representing a summary of my analysis of BitTorrent behavior under the circumstances surrounding Meta's torrenting activity. As shown below, my statistical analysis demonstrates that there is a greater than 99.99999% chance that Meta uploaded at least one piece of Plaintiffs' works to another peer. In contrast to Frederiksen-Cross's report, this analysis shows that it was exceedingly likely that Meta shared Plaintiffs' works, not exceedingly unlikely as she claims.

21. Before diving into the analysis, I will establish some facts and assumptions. First, I assume that the data in Table 4 of the Frederiksen-Cross report is correct. To ease the presentation, I copied the relevant data from the Frederiksen-Cross report into Table 1 of this report, in columns 3 and 4. Second, I assume that each torrent takes one hour to download. Given that Meta was known to be torrenting for weeks, this seems like a conservative estimate.[9] Table 2 contains the hours of downloading (or leeching) in column 2. Third, it is a fact that libtorrent uses 8 unchoke slots (i.e., unchoking is the process of uploading content to a remote peer) and that unchoke decisions are revaluated by the libtorrent client every 15 seconds. The number of times that Meta's libtorrent client evaluated unchoking decisions (and thus had an opportunity to offer a new piece to a new peer) is represented in Table 2, column 3. Last, I assume that for each peer connected to Meta's libtorrent client, there is a 50% chance[10] that the Meta libtorrent client has a piece of Plaintiffs' works already downloaded and available to share, and a 50% chance that the peer connected to Meta's libtorrent client has a piece that Meta wants.[11] In this case, the peer is unchoked due to mutual interest. For each peer at a given interval, there is thus a 25% chance of mutual interest in one piece (50% x 50%). To determine the chance of mutual interest in a piece of a torrent containing Plaintiffs' works, we multiply this 25% by the probability that a piece of a torrent contains Plaintiffs' works (Table 1, column 4). I use this to compute the chance of an unchoked peer uploading a piece containing Plaintiffs' works for each collection of torrents, and list them in column 5 of Table 2.

22. To simplify the equation below, I use "i" to represent the number of unchoke intervals, "s" to represent the number of unchoke slots, and "p" to represent the percent chance of uploading Plaintiffs' works to one peer during slot.

---

[9] See previous footnote.

[10] Why 50%? When the download starts, the Meta BitTorrent client has 0% of the pieces; when the download ends (immediately before seeding), Meta has 100% of the pieces. I'm picking the average of these extremes.

[11] The reasoning for 50% is the same as above, but for the remote peer connecting to Meta's libtorrent client.

23. Note that I use the Frederiksen-Cross report analysis directly and have not attempted to validate or reproduce those numbers. I reserve the possibility of revisiting these figures if independent analysis or additional facts change the nature or accuracy of such claims in the report.

Table 1: Overall stats for Torrents in Frederiksen-Cross report. First, third, and fourth columns are reproduced from the Frederiksen-Cross report.

Dataset	Percent Chance of Sharing Plaintiffs' Works at Least Once	# Torrents Containing Plaintiffs' Works	Percent of Pieces Containing Plaintiffs' Works
Libgen.rs Non-Fiction	72.91%	2	0.136%
Internet Archive	99.999574%	46	0.056%
Z-Library	99.99999%	146	0.023%

Table 2: Values used in statistical analysis.

Dataset	Hours Leeching	Unchoke Intervals (i)	Unchoke slots (s)	Percent chance of uploading Plaintiffs' works to one peer during slot (p)
Libgen.rs Non-Fiction	2	480	8	0.034%
Internet Archive	46	11,040	8	0.014%
Z-Library	146	35,040	8	0.00575%

24. To calculate the probability that Meta shared at least one piece of Plaintiffs' works, I focus on a Bernoulli experiment,[12] where the probability of BitTorrent picking a piece of Plaintiffs' works is fixed and statistically independent. To simplify the math, I consider the simple case of finding the probability that Meta's BitTorrent client did not share a piece of Plaintiffs' works in any of the upload slots considered during the leeching phase. This probability, P(n), is defined by:

$$P(n) = q^n$$

where q is the probability of not uploading a piece of the Plaintiffs' works during an unchoke slot and n is the number of slots considered during the leeching phase. The value of n is equal

[12] https://en.wikipedia.org/wiki/Bernoulli_trial

to the number of unchoke intervals (i) times the number of unchoke slots (s); the value of q is (1- p), where p is the percent chance of uploading Plaintiffs' works to one peer during slot.

25. The probability that Meta *did* share at least one piece of the Plaintiffs' works (which I call "U") is

$$U = 1 - P(n) = 1 - (1-p)^{(i*s)}$$

26. The value of the equation ranges from 0 to 1, corresponding to a range of a 0% chance to a 100% chance.

27. According to the first row of Table 2, the probability that Meta shared at least one piece of the Plaintiffs' works in Libgen.rs Non-Fiction is:

$$1 - (1- 0.00034)^{(480*8)} = 0.7291$$

In other words, there was a 72.91% chance that Meta shared a piece of Plaintiffs' works in Libgin.rs Non-fiction.

28. According to the second row of Table 2, the probability that Meta shared at least one piece of Plaintiffs' works in Internet Archive torrents is:

$$1 - (1- 0.0000575)^{(11040*8)} = 0.99999574$$

In other words, there was a 99.999574% chance that Meta shared a piece of Plaintiffs' works in Internet Archive torrents.

29. According to the third row of Table 2, the probability that Meta shared at least one piece of the Plaintiffs' works in Internet Archive torrents is:

$$1 - (1- 0.00014)^{(35040*8)} = 0.9999999$$

In other words, there was a 99.99999% chance that Meta shared a piece of Plaintiffs' works in Internet Archive torrents.

30. While Frederiksen-Cross claimed that such uploading of Plaintiffs' works was "exceedingly unlikely," this simple statistical analysis suggests that it was, in fact, far more likely than not. In the case of Plaintiffs' works contained in Z-Library and Internet Archive, it was nearly a foregone conclusion.

Evidence Not Provided by Meta

31. **S3 Storage Data**. Frederiksen-Cross's report does not address whether S3 storage data was reviewed, which could provide additional evidence of torrenting activities. Further, the report did not explain how and why Meta produced the list of torrented data files for the report, nor what may have been omitted from the search via this process.

32. **Evidence of likely Libgen torrenting**. The Krein report states, "Meta's script for torrenting from Anna's Archive includes an 'Example command,' which specifies a 'dataset_name' of 'libgen_rs_fic,' as well as an 'input_path' […] and an 'output_dir'", the mentioned directories are where "LibGen fiction (i.e., 'libgen_rs_fic') is the data torrented." In the Frederiksen-Cross report, she states, "In evaluating the text of comments, it is important to understand that they are discretionary annotations that a programmer can record in a script or program. They are not functional code, and they do not control any aspect of the script's actual processing."

33. However, the comment in the download_spark.py file clearly indicates torrent data associated with the Libgen.rs fiction datasets, and how they get moved to Amazon S3 storage. No justification is provided that explains why this specific, working command was in the comments, but the corresponding torrented data was not produced or discussed during interviews. Generally speaking, programmers do not provide such specific working commands as examples unless they have tested them first, i.e., they have run the command in the comment. Frederiksen-Cross's report does not speak to why this comment is there in the first place, nor whether that command to download Libgen fiction was ever executed to download that content. The evidence from interviews, command line history (see below), and S3 storage would help address this issue, but is currently missing.

34. **Command History Files**. Linux systems like the ones used to download and share Plaintiffs' works by default store a history of the commands issued by the users of those systems. Given that the torrent files were specified by command and not in the code of download_spark.py, the command history files could shed light on the specific commands executed for torrenting. These history files were not provided.

35. **Data transfer details.** The Frederiksen-Cross report references the EC2 network security configuration that blocked unsolicited inbound connections (but that did not prevent uploading in general) but does not provide evidence from billing data that data transfers to the Internet did not happen. Such evidence necessarily exists because Amazon charges its customers based on how much data is transferred out of an EC2 instance.[13] Meta should furnish this data both for periods when torrenting was not happening, and for when it was, allowing us to establish a baseline of how much data was sent from EC2 to Internet destinations before using BitTorrent, and thus establishing by comparison and with high likelihood how much data was shared with BitTorrent peers during the periods when torrenting was happening.

36. Specifically, Amazon Web Services provides their customers with a dashboard under "Billing and Cost Management" > "Cost explorer" > "New cost and usage report" where filters can be arranged to display the number of outbound GB transferred per day out of each EC2 instance. I have run this on my own EC2 account and confirmed its availability for current and historical data.

[13] https://docs.aws.amazon.com/cur/latest/userguide/cur-data-transfers-charges.html

37. **File creation/modification time data.** Information about when torrented data files were created, and when they were last modified, can identify the period when downloading was occurring. This information has not been provided.

38. **Multiple versions of code.** Frederiksen-Cross refers to multiple versions of code in paragraph 77 of her report, but only one copy of one version of code has been furnished for inspection.

39. **Libgen download details.** According to Meta_Kadrey_00107954, the following text indicates that torrenting was used for 10TB of Libgen data in 2024: "Libgen (10 TB out of 10 TB): we got almost all we want (all torrents posted after 2023-03-01) with a few ones pending." Frederiksen-Cross, however, does not mention this data at all. There needs to be corresponding details provided to explain this discrepancy.

40. **Interviews with developers**. The Frederiksen-Cross report relies on interviews from Bashlykov and Wang, among others, but the transcripts of those interviews are not provided.

Conclusion

Summary of Key Points: This rebuttal has highlighted the affirmative evidence of torrenting, the misleading statements in Frederiksen-Cross's report, and the missing evidence of the extent of Meta's leeching and seeding.

Respectfully Submitted,

David R. Choffnes, Ph.D.

February 26, 2025

APPENDIX A

CURRICULUM VITAE

David R. Choffnes

Northeastern University
Khoury College of Computer Sciences
Boston, Massachusetts 02115
United States of America

choffnes@ccs.neu.edu
https://david.choffnes.com

EDUCATION

06/2010 **Doctor of Philosophy** in Computer Science
Northwestern University, Evanston, Illinois
Thesis: *Service-Level Network Event Detection from Edge Systems*
Advisor: Fabián E. Bustamante

12/2006 **Master of Science** in Computer Science
Northwestern University, Evanston, Illinois
Advisor: Fabián E. Bustamante

05/2002 **Bachelor of Arts** in Physics and French, awarded *magna cum laude*
Amherst College, Amherst, Massachusetts

PROFESSIONAL EXPERIENCE

7/2019 – Pres. **Associate Professor**
Khoury College of Computer Sciences, Northeastern University, Boston, Mass.

7/2021 – Pres. **Executive Director, Cybersecurity and Privacy Institute**
Cybersecurity and Privacy Institute, Northeastern University, Boston, Mass.

10/2020 – 6/2021 **Security Architect**
Akamai Technologies, Cambridge, Mass.

9/2013 – 6/2019 **Assistant Professor**
College of Computer and Information Science, Northeastern University, Boston, Mass.

6/2010 – 8/2013 **Post-Doctoral Scholar (NSF/CRA Computing Innovations Fellow)**
University of Washington, Department of CSE, Seattle, Washington
Mentors: Tom Anderson and Arvind Krishnamurthy

6/2009 – 9/2009 **Research Intern**
AT&T Labs – Research, Florham Park, New Jersey

6/2007 – 9/2007 **Research Intern**
IBM T. J. Watson Research Lab, Hawthorne, New York

8/2002 – 7/2004 **Editor and author**
Deitel & Associates, Maynard, Massachusetts

HONORS AND AWARDS

2023 Best Paper (IMC 2023)
2023 Runner-up Best Student Paper (PETS 2023)
2021 ACM Senior Member
2019 Community Contribution Award (IMC 2019)
2018 DNSSEC paper (USENIX Security 2018) selected for the IRTF's Applied Networking Research Prize
2018 NSF CAREER Award
2018 Best of CCR Award for FreeBasics paper (presented at SIGCOMM 2018)
2018 Cisco Network Security Best Paper Award (NDSS 2018)
2018 QUIC paper (IMC 2017) selected for the IRTF's Applied Networking Research Prize
2017 IEEE Cybersecurity Award for Innovation
2017 USENIX Security Distinguished Paper Award
2010-2012 NSF/CRA Computing Innovations Fellowship
2010 EECS Outstanding Dissertation Award, Northwestern University

SELECTED PRESS COVERAGE

4/2023 I was quoted by Reuters in a story about Tesla vehicle cameras capturing sensitive footage of unwitting bystanders.
8/2022 Our work on identifying biases in automatic transcription services was covered by Consumer Reports.
5/2021 Our work on blocking unnecessary connections from IoT devices (`https://moniotrlab.ccis.neu.edu/pets21/`) was covered by Consumer Reports.
2/2020 Our work on systematically revealing misactivations of smart speakers and the implications of what they record (`https://moniotrlab.ccis.neu.edu/smart-speakers-study/`) appeared in the New York Times, USA Today, Vox, Fortune, and others.
9/2019 Our work on revealing users' sensitive information exposed by IoT devices was covered by The Financial Times, VICE, and others.
9/2018 Bloomberg covered our analysis of ISP throttling behavior before and after the new FCC order that permits it. Also covered by the Boston Globe, Gizmodo, Geekwire, the Verge, and other international publications such as Le Monde and Figaro.
7/2018 Coverage of our work (published at PETS 2018) on apps that spy on you. Covered by more than 200 news outlets, including Gizmodo, USA Today, WIRED, the Recode podcast (interviewed by Kara Swisher), Fox 25 Boston, Denver 7 News, The Register, and other publications worldwide. We responsibly disclosed to Google and others, they took action to mitigate this privacy risk.
1/2018 Our Wehe app for detecting net neutrality violations received substantial international attention, in large part due to Apple rejecting our app and then later reversing its decision based on public backlash. Our work was subsequently covered by dozens of news outlets, and a piece about net neutrality featuring this work appeared on VICE News' TV show.

11/2017 Article about my team's net neutrality research appeared on the PBS Newshour website, CBS News website, among others. I was also interviewed by The Takeaway (NY Times / PRI production) on a segment that aired nationally on November 21st.

11/2015 The ReCon project, which reveals and controls how personal information is leaked from mobile devices, was covered by the Boston Globe, Northeastern News, Christian Science Monitor, New Scientist, Science Codex, NBC News, MSN News, El Mundo, Le Matin, among others.

TEXTBOOKS

[1] Harvey Deitel, Paul Deitel, David Choffnes, and Chriti Kelsey. *Simply C++ : An Application-Driven Tutorial Approach*. Prentice Hall, 2005.

[2] Harvey Deitel, Paul Deitel, and David Choffnes. *Operating Systems*. Prentice Hall, Third edition, 2004.

REFEREED JOURNAL AND MAGAZINE PUBLICATIONS

[1] David Choffnes, Woodrow Hartzog, Scott Jordan, Athina Markopoulou, and Zubair Shafiq. A scientific approach to tech accountability. *Harvard Journal of Law & Technology*, 37(3), 2023.

[2] Michelle N. Meyer, John Basl, David Choffnes, Christo Wilson, and David M. J. Lazer. Enhancing the ethics of user-sourced online data collection and sharing. *Nature Computational Science*, 3(8), 2023.

[3] Arash Molavi Kakhki, Samuel Jero, David Choffnes, Alan Mislove, and Cristina Nita-Rotaru. Taking a long look at QUIC: An approach for rigorous evaluation of rapidly evolving transport protocols. *Communications of the ACM*, 62(7), 2019.

[4] Fan Zhou, Kaushik Chowdhury, and David Choffnes. Janus: A multi-TCP framework for application-aware optimization in mobile networks. *IEEE IEEE Transactions on Mobile Computing*, 2019.

[5] Liang Zhang, David Choffnes, Tudor Dumitraş, Dave Levin, Alan Mislove, Aaron Schulman, and Christo Wilson. Analysis of ssl certificate reissues and revocations in the wake of heartbleed. *Communications of the ACM*, 61(3), 2018.

[6] Quirin Scheitle, Taejoong Chung, Jens Hiller, Oliver Gasser, Johannes Naab, Roland van Rijswijk-Deij, Oliver Hohlfeld, Ralph Holz, Dave Choffnes, Alan Mislove, and Georg Carle. A First Look at Certification Authority Authorization (CAA). *ACM SIGCOMM Computer Communications Review (CCR)*, April 2018.

[7] Jing'an Xue, David Choffnes, and Jilong Wang. CDNs meet CN: An empirical study of CDN deployments in China. In *IEEE Access*, volume 5, December 2017.

[8] Rijurekha Sen, Sohaib Ahmad, Amreesh Phokeer, Zaid Ahmed Farooq, Ihsan Ayyub Qazi, David Choffnes, and Krishna P. Gummadi. Inside the walled garden: Deconstructing facebook's free basics program. *SIGCOMM CCR*, 47(5), 2017.

[9] Ashwin Rao, Arash Molavi Kakhki, Abbas Razaghpanah, Anke Li, David Choffnes, Arnaud Legout, Alan Mislove, and Phillipa Gill. Meddle: Enabling transparency and control for mobile internet traffic. *Journal of Technology Science (JoTS)*, October 2015.

[10] Arnau Gavalda, David Choffnes, John Otto, Mario Sanchez, Fabian Bustamante, Luis A. N. Amaral, Jordi Duch, and Roger Guimera. Impact of heterogeneity and socio-economic factors on massive decentralized sharing ecosystems. *Proceedings of the National Academy of Sciences (PNAS)*, October 2014.

[11] Mario A. Sánchez, John S. Otto, Zachary S. Bischof, David R. Choffnes, Fabian E. Bustamante, Balachander Krishnamurthy, and Walter Willinger. A measurement experimentation platform at the internet's edge. *IEEE/ACM Trans. Netw.*, 23(6):1944–1958, 2015.

[12] Kai Chen, David R. Choffnes, Rahul Potharaju, Yan Chen, Fabian E. Bustamante, Dan Pei, and Yao Zhao. Where the sidewalk ends: Extending the internet AS graph using traceroutes from P2P users. *IEEE Trans. Computers*, 63(4):1021–1036, 2014.

[13] Ao-Jan Su, David R. Choffnes, Aleksandar Kuzmanovic, and Fabián E. Bustamante. Drafting behind akamai: inferring network conditions based on CDN redirections. *IEEE/ACM Trans. Netw.*, 17(6):1752–1765, 2009.

[14] David R. Choffnes, Mark Astley, and Michael J. Ward. Migration policies for multi-core fair-share scheduling. *Operating Systems Review*, 42(1):92–93, 2008.

REFEREED CONFERENCE PUBLICATIONS

[1] Amogh Pradeep, Johanna Gunawan, Alvaro Feal, Woody Hartzog, and David Choffnes. Gig work at what cost?: Exploring privacy risks of gig work platform participation in the U.S. In *Proc. of PETS*, 2025 (to appear).

[2] Tianrui Hu, Daniel Dubois, and David Choffnes. IoT bricks over v6: Understanding IPv6 usage in smart homes. In *Proc. of IMC*, 2024.

[3] Daniel J. Dubois, Nicole Holliday, Kaveh Waddell, and David Choffnes. Fair or fare? Understanding automated transcription error bias in social media and videoconferencing platforms. In *Proc. of ICWSM*, 2024.

[4] Vadim Safronov, Anna Maria Mandalari, Daniel J. Dubois, David Choffnes, and Hamed Haddadi. Sunblock: Cloudless protection for iot systems. In *Passive and Active Measurement (PAM) Conference*, 2024.

[5] Aniketh Girish, Tianrui Hu, Vijay Prakash, Daniel J. Dubois, Srdjan Matic, Danny Yuxing Huang, Serge Egelman, Joel Reardon, Juan Tapiador, David Choffnes, and Narseo Vallina-Rodriguez. In the room where it happens: Characterizing local communication and threats in smart homes. In *Proc. of IMC*, 2023.

[6] Tianrui Hu, Daniel J. Dubois, and David Choffnes. Behaviot: Measuring smart home iot behavior using network-inferred behavior models. In *Proc. of IMC*, 2023.

[7] Hongying Dong, Hao Shu, Vijay Prakash, Yizhe Zhang, Muhammad Talha Paracha, David Choffnes, Santiago Torres-Arias, Danny Yuxing Huang, and Yixin Sun. Behind the scenes: Uncovering tls and server certificate practice of iot device vendors in the wild. In *Proc. of IMC*, 2023.

[8] Zeinab Shmeis, Muhammad Abdullah, Pavlos Nikolopoulos, Katerina Argyraki, David Choffnes, and Phillipa Gill. Localizing traffic differentiation. In *Proc. of IMC*, 2023.

[9] Umar Iqbal, Pouneh Nikkhah Bahrami, Rahmadi Trimananda, Hao Cui, Alexander Gamero-Garrido, Daniel J. Dubois, David Choffnes, Athina Markopoulou, Franziska Roesner, and Zubair Shafiq. Tracking, profiling, and ad targeting in the alexa echo smart speaker ecosystem. In *Proc. of IMC*, 2023.

[10] Anna Maria Mandalari, Hamed Haddadi, Daniel J. Dubois, and David Choffnes. Protected or porous: A comparative analysis of threat detection capability of iot safeguards. In *Proc. of IEEE S&P*, 2023.

[11] Monica Kowalczyk, Johanna T. Gunawan, David Choffnes, Daniel J Dubois, Woodrow Hartzog, and Christo Wilson. Understanding dark patterns in home iot devices. In *Proc. of CHI*, 2023.

[12] Amogh Pradeep, Alvaro Feal, Julien Gamba, Ashwin Rao, Martina Lindorfer, Narseo Vallina-Rodriguez, and David Choffnes. Not your average app: A large-scale privacy analysis of android browsers. In *Proc. of PETS*, 2023.

[13] Amogh Pradeep, Muhammad Talha Paracha, Protick Bhowmick, Ali Davanian, Abbas Razaghpanah, Taejoong Chung, Martina Lindorfer, Narseo Vallina-Rodriguez, Dave Levin, and David Choffnes. A comparative analysis of certificate pinning in Android & iOS. In *Proc. of IMC*, 2022.

[14] Kevin Vermeulen, Ege Gurmericliler, Italo Cunha, David Choffnes, and Ethan Katz-Bassett. Internet scale reverse traceroute. In *Proc. of IMC*, 2022.

[15] Kentrell Owens, Johanna Gunawan, David Choffnes, Pardis Emami-Naeini, Tadayoshi Kohno, and Franziska Roesner. Exploring deceptive design patterns in voice interfaces. In *Proc. of EuroUSEC*, 2022.

[16] Narmeen Shafqat, Daniel J. Dubois, David Choffnes, Aaron Schulman, Dinesh Bharadia, and Aanjhan Ranganathan. ZLeaks: Passive inference attacks on zigbee based smart homes. In *Proc. of ANCS*, 2022.

[17] Amogh Pradeep, Hira Javaid, Ryan Williams, Antoine Rault, David Choffnes, Stevens Le Blond, and Bryan Ford. Moby: A blackout-resistant anonymity network for mobile devices. In *Proc. of PETS*, 2022.

[18] Muhammad Talha Paracha, Daniel Dubois, Narseo Vallina-Rodriguez, and David Choffnes. IoTLS: Understanding TLS usage in consumer IoT devices. In *Proc. of IMC*, 2021.

[19] Johanna Gunawan, Amogh Pradeep, David Choffnes, Woodrow Hartzog, and Christo Wilson. A comparative study of dark patterns across mobile and web modalities. In *Proc. of CSCW*, 2021.

[20] Xiao Zhang, Tanmoy Sen, Zheyuan Zhang, Tim April, Balakrishnan Chandrasekaran, David Choffnes, Bruce M. Maggs, Haiying Shen, Ramesh K. Sitaraman, and Xiaowei Yang. AnyOpt: Predicting and optimizing IP anycast performance. In *Proc. of ACM SIGCOMM*, 2021.

[21] Ruimin Sun, Alejandro Mera, Long Lu, and David Choffnes. SoK: Attacks on industrial control logic and formal verification-based defenses. In *Proc. of IEEE European Security and Privacy*, 2021.

[22] Anna Maria Mandalari, Daniel J. Dubois, Roman Kolcun, Muhammad Talha Paracha, Hamed Haddadi, and David Choffnes. Blocking without Breaking: Identification and Mitigation of Non-Essential IoT Traffic. In *Proc. of the Privacy Enhancing Technologies Symposium (PETS)*, 2021.

[23] Said Jawad Saidi, Anna Maria Mandalari, Roman Kolcun, Hamed Haddadi, Daniel J. Dubois, David Choffnes, Georgios Smaragdakis, and Anja Feldmann. A Haystack Full of Needles: Scalable Detection of IoT Devices in the Wild. In *Proc. of the Internet Measurement Conference (IMC)*, 2020.

[24] Daniel Dubois, Roman Kolcun, Anna Mandalari, Muhammad Talha Paracha, David Choffnes, and Hamed Haddadi. When speakers are all ears: Characterizing misactivations of IoT smart speakers. In *Proc. of PETS*, 2020.

[25] Muhammad Talha Paracha, Balakrishnan Chandrasekara, David Choffnes, and Dave Levin. A deeper look at web content availability and consistency over HTTP/S. In *Network Traffic Measurement and Analysis Conference*, 2020.

[26] Thijs van Ede, Riccardo Bortolameotti, Andrea Continella, Jingjing Ren, Daniel J. Dubois, Martina Lindorfer, David Choffnes, Maarten van Steen, and Andreas Peter. FlowPrint: Semi-supervised mobile-app fingerprinting on encrypted network traffic. In *Proc. of NDSS*, 2020.

[27] Jingjing Ren, Daniel J. Dubois, David Choffnes, Anna Maria Mandalari, Roman Kolcun, and Hamed Haddadi. Information exposure from consumer iot devices: A multidimensional, network-informed measurement approach. *Proc. of IMC*, 2019.

[28] Taejoong Chung, Emile Aben, Tim Bruijnzeels, Balakrishnan Chandrasekaran, David Choffnes, Dave Levin, Bruce Maggs, Alan Mislove, Roland van Rijswijk-Deij, John P. Rula, and Nick Sullivan. Rpki is coming of age: A longitudinal study of rpki deployment and invalid route origins. *Proc. of IMC*, 2019.

[29] Fangfan Li, Arian Niaki, David Choffnes, Phillipa Gill, and Alan Mislove. A large-scale analysis of deployed traffic differentiation practices. In *Proc. of ACM SIGCOMM*, 2019.

[30] Taejoong Chung, Jay Lok, Balakrishnan Chandrasekaran, David Choffnes, Dave Levin, Bruce Maggs, Alan Mislove, John P. Rula, Nick Sullivan, and Christo Wilson. Is the web ready for OCSP Must Staple? In *Proc. of IMC*, 2018.

[31] Elleen Pan, Jingjing Ren, Martina Lindorfer, Christo Wilson, and David Choffnes. Panoptispy: Characterizing audio and video exfiltration from Android applications. In *Proc. of PETS*, 2018.

[32] John P. Rula, Fabian E. Bustamante, James Newman, Arash Molavi Khaki, and David Choffnes. Mile high WiFi: A first look at in-flight Internet connectivity. In *Proc. of WWW*, 2018.

[33] Jingjing Ren, Martina Lindorfer, Daniel Dubois, Ashwin Rao, David R. Choffnes, and Narseo Vallina-Rodriguez. Bug fixes, improvements, ... and privacy leaks - a longitudinal study of PII leaks across Android app versions. In *Proc. of NDSS*, 2018.

[34] Samuel Jero, Endadul Hoque, David Choffnes, Alan Mislove, and Cristina Nita-Rotaru. Automated attack discovery in TCP congestion control using a model-guided approach. In *Proc. of NDSS*, 2018.

[35] Arash Molavi Kakhki, Samuel Jero, David Choffnes, Alan Mislove, and Cristina Nita-Rotaru. Taking a long look at QUIC: An approach for rigorous evaluation of rapidly evolving transport protocols. In *Proc. of IMC*, 2017.

[36] Fangfan Li, Abbas Razaghpanah, Arash Molavi Kakhki, Arian Akhavan Niaki, David Choffnes, Phillipa Gill, and Alan Mislove. lib·erate, (n): A library for exposing (traffic-classification) rules and avoiding them efficiently. In *Proc. of IMC*, 2017.

[37] Taejoong Chung, Roland Rijswijk-Deij, David Choffnes, Alan Mislove, Christo Wilson, Dave Levin, and Bruce M. Maggs. Understanding the role of registrars in DNSSEC deployment. In *Proc. of IMC*, 2017.

[38] Brian Goodchild, Yi-Ching Chiu, Haonan Lu, Rob Hansen, Matt Calder, David Choffnes, Wyatt Lloyd, Matthew Luckie, and Ethan Katz-Bassett. The record route option is an option! In *Proc. of IMC*, 2017.

[39] Taejoong Chung, Roland van Rijswijk-Deij, Balakrishnan Chandrasekaran, David Choffnes, Dave Levin, Bruce M. Maggs, Alan Mislove, and Christo Wilson. A longitudinal, end-to-end view of the DNSSEC ecosystem. In *Proc. of USENIX Security*, 2017.

[40] James Larisch, David Choffnes, Dave Levin, Bruce M. Maggs, Alan Mislove, and Christo Wilson. CRLite: a scalable system for pushing all TLS revocations to browsers. In *Proc. of IEEE S&P*, 2017.

[41] Stevens Le Blond, Cédric Gilbert, Utkarsh Upadhyay, Manuel Gomez Rodriguez, and David Choffnes. A broad view of the ecosystem of socially engineered exploit documents. In *Proc. of NDSS*, 2017.

[42] Fangfan Li, Arash Molavi Kakhki, David Choffnes, Phillipa Gill, and Alan Mislove. Classifiers unclassified: An efficient approach to revealing ip-traffic classification rules. In *Proc. of IMC*, 2016.

[43] Christophe Leung, Jingjing Ren, David Choffnes, and Christo Wilson. Should you use the app for that? comparing the privacy implications of web- and app-based online services. In *Proc. of IMC*, 2016.

[44] Taejoong Chung, David Choffnes, and Alan Mislove. Tunneling for transparency: A large-scale analysis of end-to-end violations in the internet. In *Proc. of IMC*, 2016.

[45] Rijurekha Sen, Hasnain Ali Pirzada, Amreesh Phokeer, Zaid Ahmed Farooq, Satadal Sengupta, David Choffnes, and Krishna P. Gummadi. Inspecting the free bridge across the digital divide: Assessing the quality of facebook's free basics service. In *Proc. of IMC*, 2016.

[46] Taejoong Chung, Yabing Liu, David Choffnes, Dave Levin, Bruce Maggs, Alan Mislove, and Christo Wilson. Measuring and applying invalid SSL certificates: The silent majority. In *Proc. of IMC*, 2016.

[47] Frank Cangialosi, Taejoong Chung, David Choffnes, Dave Levin, Bruce M. Maggs, Alan Mislove, and Christo Wilson. Measurement and analysis of private key sharing in the ssl ecosystem. In *Proc. of CCS*, 2016.

[48] Jingjing Ren, Ashwin Rao, Martina Lindorfer, Arnaud Legout, and David R. Choffnes. ReCon: Revealing and controlling privacy leaks in mobile network traffic. In *Proc. of ACM MobiSys*, 2016.

[49] Arash Molavi Kakhki, Abbas Razaghpanah, Anke Li, Hyungjoon Koo, Rajesh Golani, David R. Choffnes, Phillipa Gill, and Alan Mislove. Identifying traffic differentiation in mobile networks. In *Proc. of IMC*, 2015.

[50] Yabing Liu, Will Tome, Liang Zhang, David R. Choffnes, Dave Levin, Bruce M. Maggs, Alan Mislove, Aaron Schulman, and Christo Wilson.

An end-to-end measurement of certificate revocation in the web's PKI. In *Proc. of IMC*, 2015.

[51] Ruwaifa Anwar, Haseeb Niaz, David R. Choffnes, Ítalo S. Cunha, Phillipa Gill, and Ethan Katz-Bassett. Investigating interdomain routing policies in the wild. In *Proc. of IMC*, 2015.

[52] Stevens Le Blond, David R. Choffnes, William Caldwell, Peter Druschel, and Nicholas Merritt. Herd: A scalable, traffic analysis resistant anonymity network for voip systems. In *Proc. of ACM SIGCOMM*, 2015.

[53] Ashkan Nikravesh, Hongyi Yao, Shichang Xu, David R. Choffnes, and Zhuoqing Morley Mao. Mobilyzer: An open platform for controllable mobile network measurements. In *Proc. of ACM MobiSys*, 2015.

[54] Xing Xu, Yurong Jiang, Tobias Flach, Ethan Katz-Bassett, David R. Choffnes, and Ramesh Govindan. Investigating transparent web proxies in cellular networks. In *Passive and Active Measurement (PAM) Conference*, 2015.

[55] Liang Zhang, David R. Choffnes, Dave Levin, Tudor Dumitras, Alan Mislove, Aaron Schulman, and Christo Wilson. Analysis of SSL certificate reissues and revocations in the wake of heartbleed. In *Proc. of IMC*, 2014.

[56] Kyriakos Zarifis, Tobias Flach, Srikanth Nori, David R. Choffnes, Ramesh Govindan, Ethan Katz-Bassett, Zhuoqing Morley Mao, and Matt Welsh. Diagnosing path inflation of mobile client traffic. In *Passive and Active Measurement (PAM) Conference*, 2014.

[57] Ashkan Nikravesh, David R. Choffnes, Ethan Katz-Bassett, Zhuoqing Morley Mao, and Matt Welsh. Mobile network performance from user devices: A longitudinal, multidimensional analysis. In *Passive and Active Measurement (PAM) Conference*, 2014.

[58] Umar Javed, Ítalo Cunha, David R. Choffnes, Ethan Katz-Bassett, Thomas E. Anderson, and Arvind Krishnamurthy. Poiroot: investigating the root cause of interdomain path changes. In *Proc. of ACM SIGCOMM*, pages 183–194, 2013.

[59] Stevens Le Blond, David R. Choffnes, Wenxuan Zhou, Peter Druschel, Hitesh Ballani, and Paul Francis. Towards efficient traffic-analysis resistant anonymity networks. In *Proc. of ACM SIGCOMM*, pages 303–314, 2013.

[60] Mario A. Sánchez, John S. Otto, Zachary S. Bischof, David R. Choffnes, Fabián E. Bustamante, Balachander Krishnamurthy, and Walter Willinger. Dasu: Pushing experiments to the internet's edge. In *definition*, pages 487–499, 2013.

[61] Ethan Katz-Bassett, Colin Scott, David R. Choffnes, Ítalo Cunha, Vytautas Valancius, Nick Feamster, Harsha V. Madhyastha, Thomas E. Anderson, and Arvind Krishnamurthy. LIFEGUARD: practical repair of

persistent route failures. In *Proc. of ACM SIGCOMM*, pages 395–406, 2012.

[62] John S. Otto, Mario A. Sánchez, David R. Choffnes, Fabián E. Bustamante, and Georgos Siganos. On blind mice and the elephant: understanding the network impact of a large distributed system. In *Proc. of ACM SIGCOMM*, pages 110–121, 2011.

[63] David R. Choffnes, Fabián E. Bustamante, and Zihui Ge. Crowdsourcing service-level network event monitoring. In *Proc. of ACM SIGCOMM*, pages 387–398, 2010.

[64] Kai Chen, David R. Choffnes, Rahul Potharaju, Yan Chen, Fabian E. Bustamante, Dan Pei, and Yao Zhao. Where the sidewalk ends: extending the internet as graph using traceroutes from P2P users. In *Proc. of ACM CoNEXT*, pages 217–228, 2009.

[65] David R. Choffnes and Fabian E. Bustamante. On the effectiveness of measurement reuse for performance-based detouring. In *Proc. of IEEE INFOCOM*, pages 693–701, 2009.

[66] David R. Choffnes and Fabián E. Bustamante. Taming the torrent: a practical approach to reducing cross-isp traffic in peer-to-peer systems. In *Proc. of ACM SIGCOMM*, pages 363–374, 2008.

[67] Ao-Jan Su, David R. Choffnes, Fabián E. Bustamante, and Aleksandar Kuzmanovic. Relative network positioning via CDN redirections. In *Proc. of ICDCS*, pages 377–386, 2008.

[68] Ao-Jan Su, David R. Choffnes, Aleksandar Kuzmanovic, and Fabián E. Bustamante. Drafting behind akamai (travelocity-based detouring). In *Proc. of ACM SIGCOMM*, pages 435–446, 2006.

REFEREED WORKSHOP PUBLICATIONS

[1] Johanna Gunawan, David Choffnes, Woodrow Hartzog, and Christo Wilson. Design loyalty approaches for dark patterns. In *Proceedings of 8th Workshop on Technology and Consumer Protection*, 2024.

[2] David Choffnes. A case for personal virtual networks. In *Proc. of HotNets*, 2016.

[3] Fan Zhou, Kaushik Chowdhury, and David Choffnes. Janus: Network and application-aware multi-TCP optimization engine. In *INFOCOM Poster Session*, 2016.

[4] Arash Molavi Kakhki, Fangfan Li David R. Choffnes, Alan Mislove, and Ethan Katz-Bassett. BingeOn under the microscope: Understanding T-Mobile's zero-rating implementation. In *SIGCOMM Internet-QoE Workshop*, 2016.

[5] John P. Rula, Fabian E. Bustamante, and David R. Choffnes. When IPs fly: A case for redefining airline communication. In *Proc. of HotMobile*, 2016.

[6] Ashwin Rao, Justine Sherry, Arnaud Legout, Walid Dabbout, Arvind Krishnamurthy, and David Choffnes. Meddle: Middleboxes for increased transparency and control of mobile traffic. In *Proc. of CoNEXT 2012 Student Workshop*, 2012.

[7] Ethan Katz-Bassett, David R. Choffnes, Ítalo Cunha, Colin Scott, Thomas E. Anderson, and Arvind Krishnamurthy. Machiavellian routing: improving internet availability with BGP poisoning. In *Proc. of HotNets*, page 11, 2011.

[8] Xiao Sophia Wang, David Choffnes, Patrick Gage Kelley, Ben Greenstein, and David Wetherall. Measuring and predicting web login safety. In *Proc. of ACM SIGCOMM Workshop on Measurements Up the STack (W-MUST)*, 2011.

[9] Zachary S. Bischof, John S. Otto, Mario A. Sanchez, John P. Rula, David R. Choffnes, and Fabián E. Bustamante. Crowdsourcing isp characterization to the network edge. In *Proc. of ACM SIGCOMM Workshop on Measurements Up the STack (W-MUST)*, 2011.

[10] David Wetherall, David R. Choffnes, Ben Greenstein, Seungyeop Han, Peter Hornyack, Jaeyeon Jung, Stuart E. Schechter, and Xiao Sophia Wang. Privacy revelations for web and mobile apps. In *Proc. of HotOS*, 2011.

[11] David R. Choffnes and Fabian E. Bustamante. Pitfalls for testbed evaluations of internet systems. *Computer Communication Review*, 40(2):43–50, 2010.

[12] David R. Choffnes, Jordi Duch, R. Dean Malmgren, Roger Guimerà, Fabián E. Bustamante, and Luis A. Nunes Amaral. Strange bedfellows: community identification in bittorrent. In *Proc. of IPTPS*, page 13, 2010.

[13] David R. Choffnes, Mario A. Sánchez, and Fabian E. Bustamante. Network positioning from the edge - an empirical study of the effectiveness of network positioning in P2P systems. In *Proc. of IEEE INFOCOM*, pages 291–295, 2010.

[14] David Choffnes and Fabián E. Bustamante. Exploiting emergent behavior for inter-vehicle communication. In *Proc. of Hot Topics in Autonomic Computing (HotAC)*, June 2007.

[15] David R. Choffnes and Fabián E. Bustamante. An integrated mobility and traffic model for vehicular wireless networks. In *Proc. Workshop on Vehicular Ad Hoc Networks (VANET)*, pages 69–78, 2005.

OTHER PRODUCTS

2017 Director of Technology for the *Harvest* documentary film, which appeared at the Aspen Film Festival, Toronto HotDocs Film Festival, Seattle International Film Festival, BAM Cinemafest, and Rooftop Films Summer Series. The film focuses on information gathered from my ReCon project. `https://vimeo.com/189449163`

GRANTS (ALL EXTERNAL, FUNDED)

ASPIRE Fund

2019–2020 *Shining a Light on Dark Patterns in Mobile Apps.* $75K, shared equally with Christo Wilson.

ARCEP

2017–2020 *Partnership for auditing net neutrality violations.* Contracts. $60K.

Comcast Innovation Fund

2023 *Internet Measurements to Empirically Characterize the Digital Divide.* Gift. $85K.

2018 *Personal Virtual Networks.* Gift. $71K.

2017 *Revealing and Controlling Privacy Leaks in Network Traffic.* Gift. $70K.

Consumer Reports

2022 *Understanding Profiling via Voice Assistants.* Collaborative Research. $40K.

2022 *Investigation of IoT Dark Patterns.* Collaborative Research. $40K.

2021 *Analysis of Automated Transcription Services.* Collaborative Research. $40K.

Data Transparency Lab

2015, 2018 DTL Grantee, *ReCon: Improving Transparency and Control of PII in Mobile Network Traffic.* Unrestricted gift. Primary Investigator, joint with Alan Mislove and Christo Wilson. $55K initially, $20K followup in 2018.

Department of Homeland Security Science & Technology

2017–2019 *Revealing and Controlling Privacy Leaks in Network Traffic.* Sole PI. $350K.

Google

2017 Google Research Award, *Monitoring and diagnosis of Internet QoE.* Unrestricted gift. Joint with Renata Teixeira (Inria). $62K ($31K to Northeastern).

2015 Google Research Award, *Identifying Traffic Differentiation in Mobile Networks.* Unrestricted gift. $56K.

National Science Foundation

10/24 – 09/27 CNS-2402961, *Collaborative Research: NeTS: Medium: Measurement, Modeling and Analysis to Assess Network Connectivity and Performance at Anchor Institutions.* Lead PI. Joint with Elizabeth Belding (UCSB), Alex Gamero-Garrido (UC Davis). $1.1M, $360K to Northeastern.

10/23 – 09/27 CNS-2330066, *Research Infrastructure: Mid-scale RI-1 (M1:IP): SPHERE - Security and Privacy Heterogeneous Environment for Reproducible Experimentation.* Joint with Jelena Mirkovic (lead, USC). $18M, $3.4M to Northeastern.

10/23 – 09/25 CNS-2332541, *NeTS: Continuous Monitoring and Localization of Network Neutrality Violations.* Sole PI. $99K.

10/21 – 09/26 SES-2131929, *Midscale RI: Observatory for Online Human and Platform Behavior.* Joint with David Lazer (lead, Northeastern), Christo Wilson (Northeastern). $15.7M.

10/20 – 09/25 CNS-1955227, *SaTC: Frontiers: Collaborative: Protecting Personal Data Flow on the Internet.* Lead Northeastern PI, joint with Alan Mislove, Woody Hartzog (Northeastern), Athina Markopoulou (Project Lead, UC Irvine), Zubair Shafiq (UC Davis), Konstantinos Psounis (USC). $10M (total), $1.7M (Northeastern).

10/19 – 9/22 CNS-1909020, *BehavIoT: Modeling and Controlling Internet of Things Behavior Using Network-Inferred State Machines.* Sole PI. $498K.

9/2018 – 8/2023 CNS-1750253, *CAREER: Personal Virtual Networks.* Sole PI. $513K.

10/2016–9/2019 CNS-1617728, *NeTS: Small: A Principled Approach to Enabling Policy Transparency for Mobile Networks.* co-PI, with Alan Mislove (Northeastern). $299K.

9/2016 – 8/2019 SaTC-1618955, *TWC: Small: Enabling Practical Traffic Analysis Resistance for Anonymous Communication Systems.* Sole PI. $499K.

7/2016 – 6/2020 SaTC-1564143, *TWC: Medium: Collaborative Research: Measuring and Improving the Management of Today's PKI.* co-Primary Investigator, joint with Alan Mislove, Christo Wilson (Northeastern), Dave Levin, and Tudor Dumitras (UMD). $1.2M (total), $600K (Northeastern).

1/2015 – 1/2017 CNS-1405871: *CI-New: Collaborative Research: An Open Platform for Internet Routing Experimentation.* Primary Investigator, joint with Ethan Katz-Bassett (USC) and Nick Feamster (Princeton). $1M (total), $360K (Northeastern)

9/2013 – 9/2016 CNS-1318396, *NeTS: Small: Automated Diagnosis and Root Cause Analysis of Internet Problems.* Co-Primary Investigator, joint with Arvind Krishnamurthy (University of Washington). $499K (total), $131K (Northeastern)

PwC

2019–2021 Analysis of Industrial IoT Security Threats. $100K shared with Long Lu.

Raytheon

2014 Raytheon Corporation, *Large-Scale Attacks in Multi-Level Interdependent Networks: Emerging Threats, Mitigation, and Recovery.* Grant/contract. $100K shared with Guevara Noubir, Alan Mislove, Edmund Yeh, Ravi Sundaram.

Verizon

2017–2018 *Cellular video performance measurements.* Grant/contract. $55K shared with Alan Mislove.

Amazon.com

2018 Amazon Web Services Research Grant $35K

2014 Amazon Web Services in Education Research Grants (3). $25K (in total).

ADVISING

Ph.D. Students

09/2024 – Pres. Anyu Yang (Northeastern)

08/2020 – Pres. Tianrui Hu (Northeastern)

Alumni

08/2018 – 2024 Amogh Pradeep (Northeastern, now at CrowdStrike)

08/2019 – 2024 Johanna Gunawan (Northeastern, now Professor at University of Maastricht)

08/2018 – 2023 Muhammed Talha Paracha (Northeastern, now postdoc at Bochum)

08/2015 – 2020 Fangfan Li (Northeastern, now at Facebook)

08/2014 – 2019 Jingjing Ren (Northeastern, now at Netflix)

08/2014 – 2017 Arash Molavi Kakhki (co-advised with Alan Mislove, now at Cisco/ThousandEyes)

Ph.D. Committees

Spring 2024 Narmeen Shafquat (Northeastern, committee member)

Spring 2024 Johanna Gunawan (Northeastern, **main advisor**)

Spring 2024 Amogh Pradeep (Northeastern, **main advisor**)

Fall 2023 Muhammed Talha Paracha (Northeastern, **main advisor**)

Spring 2023 Norbert Ludant (Northeastern, committee member)

Fall 2022 Bahruz Jabiyev (Northeastern, committee member)

Spring 2020 Giri Venkatadri (Northeastern, committee member)

Fall 2019 Abbas Razaghpanah (Stony Brook, committee member)

Spring 2019 Muhammad Ahmad Bashir (Northeastern, committee member)

Fall 2018 Jingjing Ren (Northeastern, **main advisor**)

Fall 2017 Jaijin Cao (Northeastern, committee member)

Spring 2017 Arash Molavi Kakhki (Northeastern, **main advisor**), currently at ThousandEyes

Fall 2016 Yabing Liu (Northeastern, committee member), currently at Twitter

Fall 2016 Liang Zhang (Northeastern, committee member), currently at Google

Spring 2015 Aldo Cassola (Northeastern, committee member), currently Professor at Universidad San Francisco de Quito (Ecuador)

SERVICE TO THE DISCIPLINE/PROFESSION

Steering Committee

2024–2026 ACM Internet Measurement Conference (IMC)

General Chair

2018 ACM Internet Measurement Conference (IMC)

Program Committee Co-Chair

2022 Internet Measurement Conference (IMC)

2019 Passive and Active Measurement (PAM) Conference
2018 IRTF Applied Networking Research Workshop (co-located with IETF 102)
2016–2018 Internet-QoE Workshop (co-located with SIGCOMM (2016-2017), ICDCS (2018))
2016 MobiData Workshop (co-located with MobiSys)
2015 All Things Cellular (co-located with SIGCOMM)

NSF Workshop Organizer

2024 Workshop on Policy-Relevant Internet Measurements
2019 Early Career Workshop
2014 Workshop on Mobile Community Measurement Infrastructure

Program Committee Member

201[4,6,7,9], 20[20-21, 23-25] ACM Internet Measurement Conference (IMC)
2017, 20[23, 25] ACM Special Interest Group on Data Communication (SIGCOMM)
2021, 2025 ACM Computer and Communications Security (CCS)
2023 Workshop on Technology and Consumer Protection (ConPro)
2022 Privacy Enhancing Technologies Symposium (PETS)
2020, 2022 IEEE Security and Privacy (Oakland)
2017, 2020 USENIX Network Systems Design & Implementation (NSDI)
2014–2017, 2019 ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT)
2018 Network and Distributed System Security Symposium (NDSS)
2014, 2017 International Conference on Mobile Systems, Applications, and Services (MobiSys) [External PC for 2017]
2016–2017 Data Transparency Lab (DTL) Grant Program
2015, 2016 International Workshop on Traffic Monitoring and Analysis (TMA)
2015 ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS)
2012, 2013 Passive and Active Measurement Conference (PAM)

Associate Editor

2015–2017 SIGCOMM Computing Communications Review (CCR)

Organization Committee

2017 CoNEXT: Travel Grant Co-Chair
2016 IMC: Works-In-Progress Chair
2016 CoNEXT: Workshop Co-Chair

Poster Committee Chair

2014 Conference of the ACM Special Interest Group on Data Communication (SIGCOMM)

Workshop Committee Member

2021 Consumer Protection (ConPro) Workshop
2013 Conference on Emerging Networking Experiments and Technologies (CoNEXT) Student Workshop
2011 SIGCOMM Workshop on Measurements "Up the Stack" (W-MUST)

2011 ACM MM '11 "Media transport and sharing"

Poster/Demo Committee Member

2021 Conference of the ACM Special Interest Group on Data Communication (SIGCOMM)

2013 Conference of the ACM Special Interest Group on Data Communication (SIGCOMM)

SOFTWARE ARTIFACTS

National Internet Observatory

Helping researchers understand online behavior.

The National Internet Observatory (NIO) aims to help researchers understand how people behave online and how platforms structure what people see. NIO is hosted at Northeastern University and is supported by a grant from the National Science Foundation. Our project consists of data collection from participants and providing access to researchers. For the former, we have developed a browser extension, Android app, and iOS app that send information back to NIO from consenting participants. For the latter, we provide qualified researcher access to such data, in an ethical, privacy-preserving way, so that researchers can study what people see online and the corresponding implications. `https://nationalinternetobservatory.org/`

Security and Privacy Heterogeneous Environment for Reproducible Experimentation (SPHERE)

Research infrastructure for at-scale, realistic and reproducible cybersecurity and privacy experimentation across diverse hardware.

To transform cybersecurity and privacy research into a highly integrated, community-wide effort, researchers need a common, rich, representative research infrastructure that meets the needs across all members of the research community, and facilitates reproducible science. This research infrastructure will offer access to an unprecedented variety of hardware, software, and other resources, all relevant to cybersecurity and privacy research, connected by user-configurable network substrate, and protected by a set of security policies uniquely aligned with cybersecurity and privacy research needs. SPHERE will offer six user portals, closely aligned with needs of different user groups, facilitating widespread adoption. It will provide built-in support for reproducibility, via easy experiment packaging, sharing, and reuse. SPHERE will build a process, a standard, and incentives for community-wide efforts to develop representative experimentation environments for cybersecurity and privacy research, and to continuously contribute high-quality research artifacts. At Northeastern, we are focusing on building a first-of-its-kind, fully remotely accessible, Internet of Things lab that can be configured and controlled without any in-person interaction. `https://sphere-project.net/`

Mon(IoT)r Lab and ReCon

Revealing and controlling privacy leaks from mobile and IoT devices.

The mobile and IoT devices that surround our everyday lives enable great new technologies, but also come a great risks to our online privacy and security. In this project, we use network traffic analysis on controlled and uncontrolled experiments with participants to infer how sensitive information can be exposed to other parties on the Internet, how dark patterns impact consumers, and what security flaws exist in IoT devices / mobile apps. We also have studied how to control the information that is exposed to other parties by blocking unnecessary network communication, and we have shared our findings with vendors as well as regulators such as the FTC when relevant. Our datasets have been downloaded several hundred times.
`https://moniotrlab.khoury.northeastern.edu/`
`https://recon.meddle.mobi`

Wehe
Identifying net neutrality violations and empowering average users to audit their providers.
Differentiation is the practice of giving different Internet service to different applications. For example, an ISP may give worse performance to YouTube (leading to rebuffering and lower quality video), but allow Netflix to stream video at full resolution without rebuffering events. This is generally considered a violation of network neutrality. We developed techniques to reliably identify when an ISP selectively gives different performance for different apps, and embedded this technology in iOS and Android apps that have seen more than 300,000 downloads and 2.5 million tests run. We make our data available to consumers, regulators, and any other interested parties. In addition, we have developed strategies to evade such differentiation and continue to undertake research to more comprehensively avoid network interference. A public comment to the FCC based on our findings was cited about a dozen times in the net neutrality rules passed in 2024. `https://wehe.meddle.mobi`

Mobilyzer *System for measuring Internet performance from mobile devices.*
Mobilyzer is an open source library for measuring network performance on mobile platforms. You can measure your network's throughput and latency, as well as other useful network metrics. Mobilyzer also supports background measurements, server-scheduled measurements, and push-based measurements. The data is collected either anonymously or from your selected account, which allows you to see your own data. The user credentials collected are not shared outside of this site, and any data used in research projects in universities are anonymized before use.
`http://www.mobilyzer-project.mobi/`

Herd *Practical, anonymous voice over IP (VOIP).*
In the face of strong adversaries with widespread surveillance, existing privacy tools fail to provide the required anonymity or performance for interactive communication (e.g., VoIP). In this work, we are building a VoIP system that resists traffic analysis under a strong adversarial model, without sacrificing

performance. We will be releasing the tool, along with source code, shortly.
`https://anonymity.ccs.neu.edu/`

SSL/PKI Security *Understanding the security of the currently deployed public key infrastructure.*
Central to the secure operation of a public key infrastructure (PKI) is the ability to revoke certificates. While much of users' security rests on this process taking place quickly, in practice, revocation typically requires a human to decide to reissue a new certificate and revoke the old one. Thus, having a proper understanding of how often systems administrators reissue and revoke certificates is crucial to understanding the integrity of a PKI. We are currently investigating how certificates are revoked, how these revocations are enforced by client software (browsers), what are the security implications of existing practices, and how we can improve the state of the art.
`https://www.sslresearch.org/`

SOFTWARE ARTIFACTS (PRIOR TO JOINING NORTHEASTERN)

Piigeon *Extension for Firefox that reveals Web login safety before you submit your password.*
Piigeon is a Firefox extension that records whether websites protect your username and password when you sign in. For most sites, the cursor will change, telling you whether your login is encrypted or if it could instead be intercepted. Over time a report of your password safety is created. Created by Xiao (Sophia) Wang.
`http://piigeon.org/`

Dasu An extension to the popular Vuze/Azureus BitTorrent client. Dasu is a dual-objective system providing ISP characterization (including the detection of network interference) and supporting Internet measurement experimentation.
Over 100,100 users as of December 2014.
`http://www.aqualab.cs.northwestern.edu/projects/Dasu.html`

SwarmScreen An extension to the popular Vuze/Azureus BitTorrent client to make it difficult to classify users' downloading behavior by looking at his/her connection patterns.
`http://www.aqualab.cs.northwestern.edu/projects/SwarmScreen.html`

NEWS A system for Network Early Warning System built by taking advantage of the natural P2P traffic. NEWS is implemented as plugin/extension for the BitTorrent Azureus client.
Over 56,000 users as of December 2014.
`http://www.aqualab.cs.northwestern.edu/projects/NEWS.html`

SideStep/DraFTP The SideStep service reuses CDN information to locate quality overlay paths in the Internet with minimum overhead. We also implemented DraFTP, and open-source FTP suite that uses SideStep to improve

download performance.
`http://www.aqualab.cs.northwestern.edu/projects/SideStep.html`

Ono A plugin/extension for the Azureus client that implements our proposed CDN-based positioning for peer selection in the popular BitTorrent system. Over 1,480,000 users as of December 2014.
`http://www.aqualab.cs.northwestern.edu/projects/Ono.html`

STRAW An integrated mobility and traffic model for Vehicular Ad-Hoc Networks (VANETs); STRAW is written for the JiST/SWANS discrete-event simulator.
`http://www.aqualab.cs.northwestern.edu/projects/STRAW.html`

SWANS++ An extension to the Jist/SWANS Discrete-event Simulator, including new/re-implementation of well-known protocols, mobility models and a steering/visualization tool.
`http://aqualab.cs.northwestern.edu/projects/swansplus2.html`

Ceratias Real-time visualization tool for the JiST/SWANS simulation platform. Also enables interaction with and online modification of the ongoing simulation, and can be detached/re-attached dynamically for performance.
`http://sourceforge.net/projects/straw/`

APPENDIX B

MATERIALS CONSIDERED

Below is a list of the materials I considered in preparing this Rebuttal Report. I also incorporate by reference all materials and sources cited in my Rebuttal Report.

Case Materials

Meta_Kadrey_00088043
Meta_Kadrey_00089791
Meta_Kadrey_00101679
Meta_Kadrey_00101699
Meta_Kadrey_00101706
Meta_Kadrey_00107954
Meta_Kadrey_00107960
Meta_Kadrey_00108151
Meta_Kadrey_00108327
Meta_Kadrey_00108336
Meta_Kadrey_00127209
Meta_Kadrey_00158789
Meta_Kadrey_00161213
Meta_Kadrey_00204223
Meta_Kadrey_00204235
Meta_Kadrey_00204707
Meta_Kadrey_00211852
Meta_Kadrey_00232764
Meta_Kadrey_00232767
Meta_Kadrey_00232876
Meta_Kadrey_00233042
Meta_Kadrey_00235448
Meta_Kadrey_00235452
Meta_Kadrey_00235739
Meta_Kadrey_00235868
META-KADREY-SC-000201-214
META-KADREY-SC-000467-472

Rule 30(b)(6) Deposition Transcript of Mike Clark, dated November 20, 2024

Litigation Filings

Plaintiffs' Motion for Leave to File Third Amended Complaint, Dkt. 395

Plaintiffs' Reply in Support of Motion for Leave to File Third Amended Complaint, Dkt. 396

Plaintiffs' Motion for Relief from Non-Dispositive Pretrial Order of Magistrate Judge, Dkt. 417

Expert Materials

Opening Report of Jonathan Krein, dated January 10, 2025

Rebuttal Report of Barbara Frederiksen-Cross, dated February 10, 2025

Barbara Frederiksen-Cross Supporting Materials